# Semantic Clustering in the Context of Legislative Amendments

**Pedro L. C. de Andrade[1] , Ellen P. R. Souza[2] , Nádia F. F. da Silva[3] , André C. P. L. F. de Carvalho[1]**

[1]Institute of Mathematical Sciences and Computation (ICMC)
University of São Paulo (USP)
Caixa Postal 668 – 13560-970 – São Carlos – SP – Brazil

[2]Rural Federal University of Pernambuco (UFRPE)
PE – Brazil

[3]Federal University of Goias (UFG)
GO – Brazil

{pedroandrade, andre}@usp.br, ellen.ramos@ufrpe.br, nadia.felix@ufg.br

***Abstract.*** *This study explores the semantic clustering of Brazilian legislative amendments through Large Language Models. Given the high number of amendments drafted annually in the Brazilian Federal Chamber of Deputies and the Federal Senate, and the consequent extensive hours spent by teams responsible for grouping these documents, the use of automated techniques for efficient analysis and organization presents itself as a beneficial alternative for legislative bodies. The study compared approaches with and without text preprocessing, varying the model's temperature parameter to assess its impact on result quality. To validate the effectiveness of the formed clusters, precision, recall, and F1 metrics were applied.*

## 1. Introduction

Among the activities carried out within the Brazilian Legislative Branch, there is the lengthy process of drafting, analyzing, and voting on Project of Bill (PL — *Projeto de Lei*) and Proposed Amendments to the Constitution (PEC — *Proposta de Emenda à Constituição*), which are key documents for updating and improving the country's legal system. Throughout this process, parliamentarians propose documents known as amendments aimed at modifying, adding to, or removing provisions from the original text of a PL or a PEC. Once created, these amendments must undergo evaluation to determine their admissibility and are subsequently debated and voted on by parliamentarians in committees and on the plenary floor. In particular, amendments addressing similar themes are grouped and discussed together, thereby increasing the complexity of the organizational work performed by the technical teams of the Federal Chamber of Deputies and the Federal Senate.

In these houses of parliament, the collection and organization of amendments is crucial to streamlining the legislative process. However, the significant number of amendments presented within short timeframes results in work overload for the teams responsible, who predominantly carry out the analysis of these documents manually. This manual execution consequently leads to limitations in agility and, even more critically, in scalability, a key challenge given the ongoing growth in the volume of projects under discussion in Brazil's federal plenaries. Thus, the adoption of automation tools aimed at accelerating the process becomes both attractive and essential. In this context, Artificial Intelligence (AI) emerges as a promising alternative for addressing the semantic clustering of legislative amendments.

In this regard, the Ulysses Project [Câmara dos Deputados 2019], an initiative established through a partnership between the Institute of Mathematics and Computer Sciences (ICMC) at the University of São Paulo (USP) and the Federal Chamber of Deputies, aimed to develop AI solutions tailored to the legislative context. This work is a product of that project and seeks to align the advances in Large Language Models (LLMs), largescale language models that employ deep learning to process and comprehend complex texts, with the task of semantic clustering. The objective is to explore the execution capabilities and performance of this emerging technology in categorizing complex documents from the Brazilian legislature.

The central objective of this work is to explore the use of OpenAI's GPT-4o and GPT-4o-mini models [OpenAI 2023] to perform the semantic clustering of legislative amendments, enabling documents with similar themes to be analyzed together, thereby reducing redundancy and making the evaluation of changes to a PL or a Proposed Amendment to the Constitution (PEC) more efficient. In this way, the feasibility of organizing amendments automatically and at scale is examined, making this approach a viable alternative to traditional methods.

Specifically, the advanced contextual understanding and language generation capabilities of these models are leveraged to form coherent groups, allowing for the performance of the clustering process to be measured. This is necessary because, although LLMs demonstrate high accuracy in various text generation tasks, they have important limitations, such as a tendency toward hallucinations — that is, the generation of incorrect information or information not grounded in the provided content. Consequently, the outputs may exhibit significant variability depending on parameters such as temperature, which controls the degree of randomness in the model's responses. Therefore, investigating how different temperature settings affect the accuracy and consistency of the clusters, with the goal of identifying configurations that strike a good balance between variability and stability, is also among the topics addressed.

Another aspect considered is the evaluation of the impact of textual preprocessing on the performance of LLMs in the semantic clustering task. To this end, two text treatment approaches will be tested: providing the full text of the amendments as input, and applying preprocessing to them. This will allow an analysis of whether text simplification helps the model capture semantic patterns or, conversely, hampers contextual analysis by removing potentially relevant terms. Precision, recall, and F1 score are the evaluation metrics used, as they measure the semantic similarity between the generated topics and the associated amendments.

Finally, the study aims to establish a knowledge base on the challenges and limitations of using LLMs in the legislative context, discussing the advantages and shortcomings of these models and suggesting directions for future work. Ultimately, the goal is to develop an application capable of performing semantic clustering as a practical solution for the organization of legislative amendments, thereby promoting greater efficiency in the analysis and categorization work carried out by parliamentary teams.

## 2. Related Works

The use of AI in legislative applications has been expanding over time, as it offers a way to design solutions that overcome the challenges posed by manual processes combined with the complexity of legal language. Additionally, LLMs represent a possibility for being applied in a way that enables legal documents to be interpreted and categorized at scale.

### 2.1. The Use of LLMs in the Legal Domain

Solutions for the legislative context based on LLM have been explored in recent work [Vayadande et al. 2024], where, in addition to automatically generating texts, these models can also identify inconsistencies and perform tasks such as sentiment analysis and Named Entity Recognition (NER) with high precision. Using the GPT-3.5 model [OpenAI 2023] and combining various preprocessing techniques, text normalization, and stopword removal, these systems are capable of detecting anomalies and facilitating the understanding of complex and extensive documents. In tests, they achieved higher accuracy rates and lower error compared to the manual process and software without the use of LLM.

### 2.2. Information Retrieval in the Legal Domain

In a Brazilian context, there are initiatives aimed at improving the efficiency of accessing and analyzing documents through AI. The Ulysses project [Souza et al. 2021] reflects this. Focused on the Federal Chamber of Deputies, there is an information retrieval pipeline designed to increase transparency and provide support for legislative activities. To achieve this, the developed system uses variants of the BM25 algorithm, which retrieves relevant documents in response to specific queries, focusing on legislative proposals such as Bills and Proposed Amendments to the Constitution. Furthermore, the approach involves the removal of stopwords, the use of stemming, and combinations of unigrams and bigrams to optimize the relevance of the results. This pipeline is essential for preliminary consultations conducted by Chamber technical staff, who need to identify redundant proposals or those related to existing ones, avoiding duplicate efforts [Souza et al. 2021].

### 2.3. Clustering of Legal Amendments

#### 2.3.1. Brazilian Senate

In addition to information retrieval initiatives in the Chamber of Deputies, other important research fronts in Brazil have explored the organization and analysis of legislative amendments through advanced NLP techniques. In the context of the Federal Senate, different approaches have been evaluated to group similar amendments with a focus on improving the efficiency of the legislative process [Pressato et al. 2024]. In the study, the BM25 techniques and variations were compared with the SBERT [Reimers and Gurevych 2019] in the process of evaluating similarity between amendments based on their textual content in three scenarios: full text, only the justification section, and only the main text. The results showed better performance, in terms of recall, of BM25L, with exact matches occurring when the full text of the amendments was considered. Moreover, the use of the stemming algorithm RLSP [Moreira and Huyck 2001] proved essential in increasing the precision of the results [Pressato et al. 2024].

#### 2.3.2. Italian Senate

Techniques for clustering legislative amendments from the Italian Senate have been explored with the goal of improving management and facilitating simultaneous voting processes [Agnoloni et al. 2022]. The main challenge in the study was identifying groups of amendments with similar textual formulations, sometimes scattered throughout a dossier, but proposing similar changes in different parts of a law. The developed system used lexical similarity metrics to group almost duplicate texts, prioritizing textual proximity over complete semantic similarity. This approach allowed the Italian Senate to address the phenomenon of legislative obstruction, where opposition parties submit a large number of similar amendments with small variations in order to delay the legislative process. The algorithm was integrated as an experimental feature in the internal amendment management system, providing an interface for the detection, visualization, and navigation between clusters of similar amendments. This study highlights the importance of extracting and normalizing domain-specific features, such as legislative citations within texts, to improve the effectiveness of automatic clustering. Thus, the approach emphasizes that the creation of clusters does not require prior information or manually annotated datasets, adopting an unsupervised model to handle variation in legislative texts.

## 3. Methodology

### 3.1. Dataset

The dataset available at [dos Deputados 2024], which serves as the basis for this project, consists of a total of 6,320 legislative amendments from the Brazilian Federal Senate, covering a 13-year period from 2010 to 2022. The amendments are distributed as shown in Table 1.

In addition to being organized by year, the amendments are grouped by legislative proposal, indicating the specific PL or PEC to which they refer. The amendments vary in length and complexity, containing both full texts and explanatory sections, such as justifications and technical excerpts. Although it was not possible to apply the developed program to all amendments, amendments related to Article 21 of PL 280 from 2020 (71 documents) and Article 4 of PL 8167 from 2022 (22 documents) were made available.

For these two sets of amendments, it was calculated using an encoder based on OpenAI models that the amendments related to Article 21 of PL 280/2020 collectively amount to 8,561 tokens, while those related to Article 4 of PL 8167/2022 total 6,180 tokens.

### 3.2. Pipeline

To develop the proposed project, the pipeline presented in Figure 1 was adopted.

| Year | Number of Amendments |
|---|---|
| 2010 | 252 |
| 2011 | 417 |
| 2012 | 949 |
| 2013 | 811 |
| 2014 | 420 |
| 2015 | 415 |
| 2016 | 533 |
| 2017 | 291 |
| 2018 | 461 |
| 2019 | 326 |
| 2020 | 546 |
| 2021 | 522 |
| 2022 | 377 |

**Table 1. Distribution of Amendments per Year in the Dataset**

In order to measure the impact of text preprocessing on clustering, two different approaches were adopted: submitting the amendment text without any preprocessing; or removing stopwords and normalizing the content.

After textual processing, the amendments are passed to the LLM, which is instructed to perform semantic clustering, where each group should be assigned a brief topic description and each amendment may belong to only one group. This process is repeated for five different temperature values (0, 0.25, 0.50, 0.75, and 1) in order to assess the impact of this parameter on execution, since increasing the temperature (ranging from 0 to 1) raises the degree of randomness in the LLM's responses.

To validate the quality of the clusters, the semantic similarity between the amendments and the topics generated by the model for each group was measured. The evaluation is carried out using three main metrics:

- Precision: Measures the proportion of correctly grouped amendments within a group, that is, how many amendments assigned to a group actually belong to it.
- Recall: Assesses the model's ability to identify all relevant amendments for a given group. In other words, it evaluates the completeness of the clustering in relation to the amendments that should be included.
- F1-Score: Combines precision and recall into a harmonic mean, serving as a metric that balances both aspects. A high F1-score indicates that the model is both precise and comprehensive at the same time.

Using these metrics for each group, the averages are calculated to obtain the precision, recall, and F1-score of the execution based on a given workflow and temperature value. This process is repeated five times, enabling a statistical analysis of the results.

### 3.2.1. Text Preprocessing

From the file containing the amendments provided as input, if the pre-processing described earlier is applied, the NLTK (*Natural Language Toolkit*) [Bird and Loper 2004]

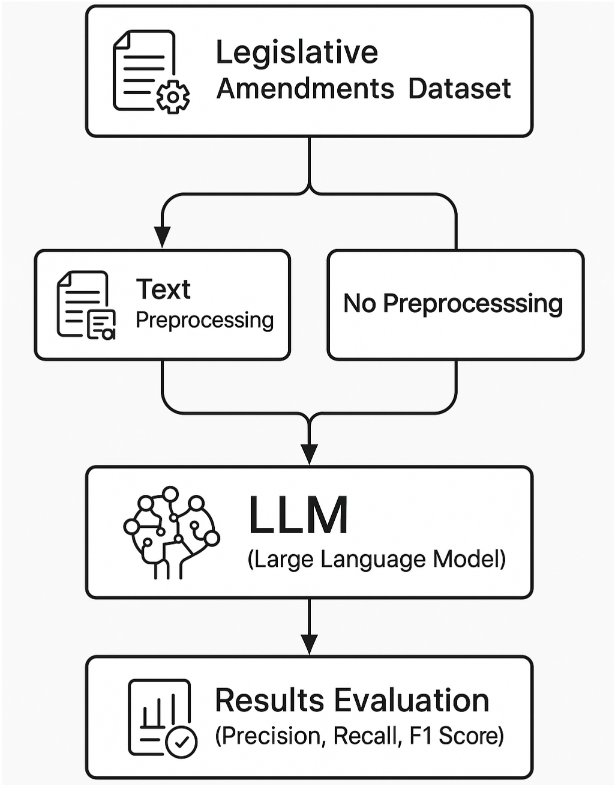


**Figure 1. Project Pipeline**

library is used to remove stopwords from a list for Portuguese.

#### 3.2.2. Use of LLMs

Access to the GPT-4o and GPT-4o Mini models involves using the API (Application Programming Interface) from OpenAI to interact with the GPT-4o and GPT-4o-mini models, both of which have a context window of 128 thousand tokens. Additionally, the temperatures used in model executions are explained, varying from 0.0 to 1.0, and how this variation impacts the generation of groups. During the implementation process, the requirement for a sophisticated prompt was evident, necessitating the listing of certain constraints to ensure that the integrity of the semantic grouping process, as well as the desired output format, were respected. This can be seen in Table 2. This prompt, within the encoder standards of OpenAI models, totals 170 tokens; thus, the execution based on the amendments related to Art. 21 of PL 280 of 2020 had an input of 8337 tokens, while the execution based on the amendments related to Art. 4 of PL 8167 of 2022 had an input of 6350 tokens.

#### 3.2.3. Results Evaluation

The evaluation of the semantic groupings generated by LLMs was conducted using BERTScore [Zhang et al. 2020], an automatic metric widely employed for assessing generated text tasks. Leveraging semantic and contextual similarity, it differs from traditional

| Function | Prompt |
|---|---|
| Semantic Clustering | You are an AI assistant specialized in the legislative sector and your task is to semantically group the parliamentary amendments provided by the user and generate groups containing the topic of each group and the amendments in it.<br><Restrictions><br>1. ALWAYS generate the groups in Portuguese.<br>2. Each amendment must appear in ONLY one group. Only include the amendments, DO NOT explain why it is in the group.<br>3. There MUST NOT be any group without amendments.<br>4. DO NOT make up amendments, use only the ones given to you along with their respective IDs.<br>5. Generate each group topic succinctly, using at most one sentence.<br>6. The output must STRICTLY follow the format:<br>{format_instructions}<br></Restrictions><br>The amendments are listed below:<br>{amendments} |

**Table 2. Adopted Prompts**

metrics that rely solely on exact word matches. The similarity score is calculated between two text sequences by comparing each token in the candidate text, which represents the generated group's topic, with each token in the reference text, which includes the amendments belonging to that group. It is important to note that evaluation using BERTScore does not guarantee precise assessment; it is merely one of the available metrics.

However, instead of performing an exact match between words, contextual embeddings, vectors that capture the meaning and context of each word, are utilized, generated by a pre-trained model based on BERT (Bidirectional Encoder Representations from Transformers) [Devlin et al. 2019]. Subsequently, for each token in the candidate text, BERTScore identifies the most similar token in the reference text by calculating the cosine distance to measure the proximity between the vectors of both tokens. Thus, it is possible to capture correspondences between words that, while different in form, share similar meanings [Zhang et al. 2020].

## 4. Results

The results and a detailed analysis of the application of the GPT-4o and GPT-4o-mini models in the semantic clustering of selected legislative amendments from the dataset are presented. It is noteworthy that for each of the sets, both text processing approaches were applied across all temperature variations, with each process being repeated five times in order to collect data for conducting a statistical analysis of the results. The results obtained from the applied metrics can be seen in Tables 3, 4, 5, 6, 7, and 8.

### 4.1. Without Text Preprocessing

It was found that the GPT-4o model outperformed the GPT-4o-mini for the adopted metrics. From this, it can be inferred that the more robust architecture, combined with the greater computational capacity of the GPT-4o, allows for a more efficient processing of textual data, resulting in a more accurate semantic analysis. It is noteworthy that for the GPT-4o, lower temperature values produced better results, indicating that this model is more optimized for more controlled temperatures.

### 4.2. With Preprocessing

It was observed once again that the GPT-4o generated more precise and concise groups compared to the GPT-4o-mini when applying the preprocessing step to the amendments, especially at lower temperatures. Furthermore, as this parameter increased, the performance of the first model tended to decrease, reinforcing that it is inclined to generate less consistent groupings with a greater tolerance for randomness in responses.

### 4.3. Comparison between Approaches

In general, the precision obtained for the preprocessed amendments was lower than that achieved with unprocessed amendments. For the GPT-4o, there was approximately a 10% decrease in precision, while for the GPT-4o-mini, the metric was about 1% lower. This is evidenced in Tables 3 and 4 and shows that, although preprocessing simplifies the text by removing connecting terms and other low-semantic words, the LLM performs better when dealing with the complete text, achieving more accurate interpretations and forming more refined groups.

Based on the recall values presented in Tables 5 and 6, for both PL 280/2020 and PL 8167/2022, the recall values for preprocessed amendments are generally lower than those for their unprocessed counterparts. Although GPT-4o-mini exhibits less variance, its recall with preprocessing also remains slightly lower or comparable, confirming that full-text input allows the models to retrieve more relevant groupings. This supports the idea that preserving the original structure and semantics of the amendments contributes to better clustering performance.

The F1-score results shown in Tables 7 and 8 reinforce the pattern observed in both precision and recall metrics: models tend to perform better when processing the original, unaltered text of the amendments. In both tables, the highest F1-score is obtained by GPT-4o without preprocessing, also GPT-4o-mini follows the same behavior. These results highlight that while preprocessing may simplify the input, it also removes important contextual or semantic cues, ultimately leading to less effective clustering by large language models.

## 5. Limitations

Although the main objective of the project was achieved, proving that it is possible to perform semantic clustering of legislative amendments using LLMs, the conduct of this work was not without its challenges and limitations. Below are the most impactful ones on the final result.

**Table 3. Clustering Precision for PL 280-2020**

| Temperature | GPT-4o | GPT-4o-mini | GPT-4o [Preproc.] | GPT-4o-mini [Preproc.] |
|---|---|---|---|---|
| 0.0 | **0.7789** | 0.6960 | 0.7085 | 0.6846 |
| 0.25 | **0.8000** | 0.7031 | 0.7178 | 0.6883 |
| 0.5 | **0.7854** | 0.7043 | 0.7201 | 0.6787 |
| 0.75 | **0.7671** | 0.7105 | 0.7109 | 0.6865 |
| 1.0 | **0.7664** | 0.7235 | 0.6937 | 0.6727 |

**Table 4. Clustering Precision for PL 8167-2022**

| Temperature | GPT-4o | GPT-4o-mini | GPT-4o [Preproc.] | GPT-4o-mini [Preproc.] |
|---|---|---|---|---|
| 0.0 | **0.8007** | 0.7660 | 0.7223 | 0.6960 |
| 0.25 | **0.7857** | 0.7518 | 0.7179 | 0.7032 |
| 0.5 | **0.7756** | 0.7308 | 0.7067 | 0.7043 |
| 0.75 | **0.7517** | 0.7494 | 0.6858 | 0.7105 |
| 1.0 | **0.7614** | 0.7190 | 0.6984 | 0.7235 |

**Table 5. Clustering Recall for PL 280-2020**

| Temperature | GPT-4o | GPT-4o-mini | GPT-4o [Preproc.] | GPT-4o-mini [Preproc.] |
|---|---|---|---|---|
| 0.0 | **0.6248** | 0.5869 | 0.6000 | 0.5727 |
| 0.25 | **0.6101** | 0.5923 | 0.5978 | 0.5724 |
| 0.5 | **0.6135** | 0.5839 | 0.5986 | 0.5679 |
| 0.75 | **0.6079** | 0.6042 | 0.5992 | 0.5734 |
| 1.0 | **0.6170** | 0.6067 | 0.5887 | 0.5744 |

**Table 6. Clustering Recall for PL 8167-2022**

| Temperature | GPT-4o | GPT-4o-mini | GPT-4o [Preproc.] | GPT-4o-mini [Preproc.] |
|---|---|---|---|---|
| 0.0 | **0.6119** | 0.5665 | 0.5520 | 0.5727 |
| 0.25 | **0.5753** | 0.5671 | 0.5489 | 0.5723 |
| 0.5 | **0.5744** | 0.5606 | 0.5655 | 0.5679 |
| 0.75 | **0.5744** | 0.5573 | 0.5516 | 0.5734 |
| 1.0 | 0.5568 | 0.5684 | 0.5431 | **0.5744** |

**Table 7. Clustering F1-Score for PL 280-2020**

| Temperature | GPT-4o | GPT-4o-mini | GPT-4o [Preproc.] | GPT-4o-mini [Preproc.] |
|---|---|---|---|---|
| 0.0 | **0.6880** | 0.6340 | 0.6527 | 0.6162 |
| 0.25 | **0.6869** | 0.6344 | 0.6474 | 0.6208 |
| 0.5 | **0.6835** | 0.6487 | 0.6511 | 0.6151 |
| 0.75 | **0.6777** | 0.6487 | 0.6472 | 0.6194 |
| 1.0 | **0.6819** | 0.6341 | 0.6432 | 0.6209 |

### 5.1. Hardware and API Costs

Initially, the intention was to adopt LLMs that could be installed and run locally completely free of charge. However, to accomplish this efficiently, the hardware neces-

**Table 8. Clustering F1-Score for PL 8167-2022**

| Temperature | GPT-4o | GPT-4o-mini | GPT-4o [Preproc.] | GPT-4o-mini [Preproc.] |
|---|---|---|---|---|
| 0.0 | **0.6825** | 0.6375 | 0.6216 | 0.5821 |
| 0.25 | **0.6592** | 0.6428 | 0.6172 | 0.5892 |
| 0.5 | **0.6498** | 0.6330 | 0.6277 | 0.6168 |
| 0.75 | **0.6494** | 0.6272 | 0.6124 | 0.5991 |
| 1.0 | 0.6256 | **0.6424** | 0.5994 | 0.5895 |

sary to accommodate the complexity and size of these models is robust, requiring high-performance GPU (Graphic Processing Unit) with substantial memory capacity, which is financially unfeasible for this work. Alternatively, the use of APIs from an external provider, OpenAI, was adopted, and although this incurred a lower cost, it was still a significant factor, which is why only two sets of amendments were grouped and each grouping was repeated only five times.

### 5.2. Risk of Hallucination

The risk of hallucination, inherent in the use of LLMs, also had to be considered, given that, in the legislative context, it is essential that there are no inaccurate groupings. To mitigate this issue, prompts with restrictions were adopted that warned against repeating amendments in more than one group, ensuring that all amendments were grouped and prohibiting the use of an amendment that was not among those provided, that is, avoiding hallucinations to the extent of creating a non-existent amendment. Additionally, a critical assessment was carried out throughout the process to identify potential hallucinations. Despite all the approaches taken to reduce the likelihood of hallucination, this risk cannot be reduced to zero; therefore, the results must be monitored to identify the occurrence of this phenomenon.

## 6. Conclusion

The analyses conducted allowed for an understanding of the performance of the models and the influence of the temperature parameter and the adoption of pre-processing on the amendments. It was concluded that the GPT-4o outperformed the GPT-4o-mini in precision and recall metrics, although it is still far from achieving state-of-the-art performance. However, it is important to emphasize that, despite providing a quantitative assessment, the numerical results obtained do not guarantee definitive conclusions, as statistical significance was not reached due to the small size of the corpus used, which consisted of only two sets of amendments. The lack of a more robust database undermines the reliability of the results, which may not reflect consistent or generalizable patterns. Thus, the quantitative and statistical evaluations performed require a significance analysis to validate the conclusions, allowing for a more robust and precise comparison between the analyzed models.

## References


Agnoloni, T., Marchetti, C., Battistoni, R., and Briotti, G. (2022). Clustering similar amendments at the Italian senate. In *Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference*, pages 39–46, Marseille, France. European Language Resources Association.

Bird, S. and Loper, E. (2004). NLTK: The natural language toolkit. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 214–217, Barcelona, Spain. Association for Computational Linguistics.

Câmara dos Deputados (2019). Câmara lança Ulysses, robô digital para facilitar acesso a informações legislativas. [Acessado em: 15 de julho de 2024].

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

dos Deputados, C. (2024). u4 ordenado completo. Acessado em: 17 dez. 2024.

Moreira, V. and Huyck, C. (2001). A stemming algorithm for the portuguese language. In *Proceedings of the Eighth International Symposium on String Processing and Information Retrieval (SPIRE 2001)*, Porto, Portugal. IEEE.

OpenAI (2023). Gpt-4 technical report. Acessado em 4 de março de 2024.

Pressato, D., de Andrade, P. L. C., Junior, F. R., Siqueira, F. A., Souza, E. P. R., da Silva, N. F. F., de Souza Dias, M., and de Leon Ferreira de Carvalho, A. C. P. (2024). Natural language processing application in legislative activity: a case study of similar amendments in the Brazilian senate. In Gamallo, P., Claro, D., Teixeira, A., Real, L., Garcia, M., Oliveira, H. G., and Amaro, R., editors, *Proceedings of the 16th International Conference on Computational Processing of Portuguese - Vol. 1*, pages 614–619, Santiago de Compostela, Galicia/Spain. Association for Computational Lingustics.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Souza, E., Vitório, D., Moriyama, G., Santos, L., Martins, L., Souza, M., Fonseca, M., Félix, N., de Carvalho, A. C. P. d. L. F., Albuquerque, H. O., and Oliveira, A. L. I. (2021). An information retrieval pipeline for legislative documents from the brazilian chamber of deputies. In *Legal Knowledge and Information Systems*, pages 119–126. IOS Press.

Vayadande, K., Bhat, A., Bachhav, P., Bhoyar, A., Charoliya, Z., and Chavan, A. (2024). Ai-powered legal documentation assistant. In *Proceedings of the 4th International Conference on Pervasive Computing and Social Networking (ICPCSN)*. IEEE.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2020). BERTSCORE: Evaluating text generation with BERT. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*. Cornell University and ASAPP Inc.